Beyond the Frame
Building a Virtual Poster Wall with Linked Open Data

Linking Cinema Context with the EYE film poster collection

Leon van Wissen, l.vanwissen@uva.nl
FIAF Cataloguing and Documentation Commission Workshop - 25 May 2023, Stockholm
Introduction

Leon van Wissen
Data Engineer, University of Amsterdam
l.vanwissen@uva.nl

https://www.create.humanities.uva.nl/
Cinema Context

About

- Encyclopedia of Dutch film culture
- History of cinema exhibition and distribution
  - ‘New Cinema History’: moviegoing as a social and cultural phenomenon

Content

- Films, Venues, Screenings, Persons, Companies
- Outgoing links to other datasets (IMDb)

Statistics

- 108k+ screenings
- 45k+ films
- 1.6k+ cinemas

Cinema Context in RDF

**Linked Open Data (LOD)**
Gaining momentum in digital cultural heritage

- Each *thing* has a URI as unique identifier
  - We use its permalink (e.g. [http://www.cinemacontext.nl/id/F020802](http://www.cinemacontext.nl/id/F020802))
- Statements about these *things* are stored in 'triples' in a graph database
- External relations

With these URIs we (and others!) can use the web to describe our data

---

Cinema Context

by Cinema Context

Created 9 months ago
3,442,725 statements

Cinema Context is an online film encyclopaedia with more than 100,000 film screenings since 1895. It provides insight into the DNA of Dutch film and cinema culture and is praised by film historians worldwide. With a DANS Small Data Project grant, this dataset has been converted to a Linked Data format (RDF). For more information, see: https://uvacreate.gitlab.io/cinema-context/cinema-context.rdf/

Graphs
- default
- metadata

Example resources
- Tuschinski
- Jean C.F.Th. Desmet
- Jonge harten (1936)

Dependent queries
- Cinemas per city
- Programmes per city
- 20 most popular films per city (variable)

Dependent stories
- Introduction to Cinema Context

Hosted by the UvA University Library: https://lod.uba.uva.nl/Cinema-Context/Cinema-Context
Linked Data on the web

Our specialised Cinema Context dataset can add to this:

- Programmes and Events
- Reviews and censorships
- Companies and Theaters
Idea for a hackathon

Can we create a virtual poster wall?
A visual source: The Memory

The Memory Database (https://geheugen.delpher.nl/)
Online heritage collections – paintings, drawings, stamps, posters and photographs

- EYE Museum Collection
  - Poster Project: https://geheugen.delpher.nl/en/geheugen/pages/collectie/Afficheproject/Collectie+EYE
  - But, not yet Linked Open Data...

"The earliest poster dates back to 1896 and the collection is still being expanded on a daily basis with contemporary material. The 6000 posters on the website The Memory constitute the main part of the film posters designed in the Netherlands."
Step 1: Obtain the data from The Memory

- Search on "Eye AND afficheproject"
  - https://geheugen.delpher.nl/nl/geheugen/results?query=Eye+AND+afficheproject&maxperpage=36&coll=ngvn&page=1
- Scrape/harvest the page's source
  - Get all meta-data for every tile

https://gitlab.com/uvacreate/cinema-context/hackalod-2020
JSON data example of a single resource in The Memory

```
"urn:gvn:NAGO02:EYE-A33169": {
  "title": "Tartuffe",
  "alternative": "Tartuffe",
  "creator": "Frans Bosen",
  "objectlevelId": "NAGO02:EYE-A33169",
  "thumbnail": "https://imageviewer.kb.nl/ImagingService/imagingService?userresolver=false&id=http%3A%2F%2Fresolver.kb.nl%2Fresolve%3Furn%3Agvn%3ANAGO02%3AEYE-A33169%26size%3Dlarge&maxw=294",
  "collectionStringNL": "Afficheproject",
  "collectionStringEN": "Poster project",
  "created": false,
  "originalStartDate": "Fri Jan 01 00:19:32 CET 1926",
  "startDate": "1926.01.01"
}
```

URI

https://resolver.kb.nl/resolve?urn=urn:gvn:NAGO02:EYE-A33169
Step 2: Transform the data to RDF

**JSON**

```
"urn:gvn:NAGO02:EYE-A33169": {
  "title": "Tartuffe",
  "creator": "Frans Bosen",
  "startDate": "1926.01.01"
}
```

**Turtle**

```
<https://resolver.kb.nl/resolve?urn=urn:gvn:NAGO02:EYE-A33169> a schema:Poster ;
  schema:name "Tartuffe" ;
  schema:creator "Frans Bosen" ;
  schema:dateCreated "1926"^^xsd:year ;
  schema:additionalType <http://vocab.getty.edu/aat/300027221> .
```
Once loaded in our own *triplestore*, let's take a look at the Poster dataset:

- **Poster query**
- **Poster gallery query**
Step 3: Link the data

We now have two independent datasets. **These are not yet linked!**

We can do so with a **linkset** with the property:

- schema:about

By matching on:

- Name
- Date

https://www.imdb.com/title/tt0017448

"1926"^^xsd:gYear

"Tartuffe"^^xsd:string
**Matching**

- **Pre-processing:**
  - Removing determiners
  - Removing parenthesis

- **Very certain:**
  - Both name + date are exactly the same

- **Somewhat certain:**
  - Name + date have a fuzzy match

- **Manual check:**
  - Name is not unique and gives multiple options
  - Date is too far apart

<table>
<thead>
<tr>
<th>poster_uri</th>
<th>poster_name</th>
<th>poster_date</th>
<th>diff_days</th>
<th>cc_date</th>
<th>cc_names</th>
<th>cc_uri</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://resolver.kb.nl/resolve?urn=urn:gn:NAGO02:EYE-A33155">https://resolver.kb.nl/resolve?urn=urn:gn:NAGO02:EYE-A33155</a></td>
<td>Als ik koning was</td>
<td>1930-01-01</td>
<td>-1461</td>
<td>1934-01-01</td>
<td>['Wenn ich König wär', 'Als ik Koning was']</td>
<td><a href="http://www.cinemacontext.nl/id/F002281">http://www.cinemacontext.nl/id/F002281</a></td>
</tr>
</tbody>
</table>
Building a poster wall

Website with user input (e.g. as filter or parameter in your query):

- Date
- Cinema (optional)

Query:

- Give me every instance of type `schema:Movie` that is screened on a `xsd:date` in a `schema:MovieTheater`.
- Give me the `schema:image` value (thumbnail) of every instance of type `schema:Poster` that is about that same `schema:Movie` instance.

See query: [Query](#)
Result: https://www.hicsuntleones.nl/plakmuur/eye/
See also: https://uvacreate.gitlab.io/cinema-context/cinema-context-rdf/events/hackalod2020/

Thanks to Thunnis van Oort (RU) and Menno den Engelse (http://islandsofmeaning.nl/) for building this together.
Advice

Open Data

● Provide URIs and pids to any object in your data
● Provide endpoints and APIs
  ○ IIIF for images
  ○ SPARQL for RDF
● Publish your data under a permissive license (e.g. CC licenses or PD)

Sharing / working together

● Team up with other archives and researchers!
Example of LOD: what are the possibilities?
Bringing data together in the same model and querying multiple datasets (incl. Wikidata) at once: https://uvacreate.gitlab.io/cinema-context/cinema-context-lod-workshop-homer-2022/

Contact

Leon van Wissen
University of Amsterdam
l.vanwissen@uva.nl